ISSN 1561-5359. Штучний інтелект, 2016, № 2

УДК 32.973.202:07.681

О.М. Марченко-Бабіч
Військовий інститут Київського національного університету імені Тараса Шевченка, Україна
вул. М.Ломоносова, 81, м.Київ, 03680

УДОСКОНАЛЕННЯ ПІДХОДІВ ДО МОНІТОРИНГУ ІНШОМОВНИХ ЗАСОБІВ МАСОВОЇ ІНФОРМАЦІЇ ЩОДО ПОДІЙ В УКРАЇНІ

O.M.Marchenko-Babich
The Military Institute of Taras Shevchenko National University of Kyiv, Ukraine
Lomonosova str., 81, Kyiv, 03680

THE IMPROVEMENT OF APPROACHES TO FOREIGN MASS MEDIA MONITORING ON THE SUBJECT OF THE SITUATION IN UKRAINE

У статті розглянуто шляхи удосконалення підходів до моніторингу іншомовних засобів масової інформації щодо подій в Україні, з урахуванням особливостей соціально-політичної обстановки, яка склалась в Україні останнім часом. Наведено основні результати аналізу провідних світових ЗМІ та запропоновано найбільш відповідні інструменти технології Сентимент-аналізу для опрацювання інформаційних потоків.

Ключові слова: інформаційний простір, моніторинг, дискурс, лінгвістичний, Сентимент-аналіз, класифікація.

The article studies the ways to improve the approaches to monitoring of foreign mass media on the subject of the situation in Ukraine regarding the social and political environment features. The main results of analysis of the key world Mass media are listed. The most adequate tools of the Sentiment-analysis for the information streams are proposed.

Key words: information environement, monitoring, discourse, linguistic, Sentiment-analysis, classification.

Вступ

Сучасний інформаційний простір наповнений повідомленнями іншомовних засобів масової інформації (ЗМІ), що висвітлюють події різних країн, у тому числі й України.

Опрацювання та аналіз іншомовних інформаційних повідомлень вимагає значних витрат часу, ускладнюючи процес виявлення інформаційних джерел та своєчасне реагування. Великий обсяг повідомлень іншомовних ЗМІ та стислість термінів викликають необхідність удосконалення підходів до моніторингу інформаційного простору [1], що включає урахування особливостей цільових груп та інших важливих факторів. Удосконалення підходів до моніторингу іншомовних ЗМІ потребує використання програмних систем для об'єктивізації добування інформації зі ЗМІ та перевірки її достовірності.

Аналіз останніх досліджень

У ході досліджень за даним напрямом з'ясувалося, що останнім часом велика приділяється увага удосконаленню засобів оперативного й поглибленого аналізу іншомовних інформаційних масивів, у тому числі текстових повідомлень ЗМІ. Це може досягатись шляхом автоматизації процесу багатовимірної аналітичної обробки повідомлень ЗМІ, що включає інтеграцію процесу обробки даних та динамічної актуалізації вихідних умов з відповідних інформаційних джерел. Для аналізу інформаційного контенту і прогнозування його розвитку в Інтернет-просторі, у [2], [3] запропоновано інструментарій, що поєднує пошук релевантних джерел, аналіз вибраного контенту, прогноз його розвитку і складається з математичних методик та технологічних компонувань даних у єдиний профіль для конкретної галузі за напрямом

застосування. Про зростаючу важливість моніторингу та контролю відповідними державними структурами різних сфер суспільно-політичного життя свідчить і те, що на даний час відбувається розвиток системи виявлення терористичної загрози в Інтернет-комунікаціях з боку деяких країн, в інтересах структур безпеки Європейського союзу [4].

Проте, незважаючи на значний науковий доробок у даній сфері, аналіз проблемних питань у методичних підходах вивчення іншомовних ЗМІ [5] свідчить, що однією з найслабших ланок є процедура побудови формальної моделі їх семантики. Також, недостатня увага приділяється розробці програмно-математичних засобів опрацювання зарубіжних іншомовних ЗМІ з приділенням уваги лінгвістичній складовій цього процесу.

Тому, **метою і основним змістом статті є** визначення основних шляхів з удосконалення підходів до моніторингу іншомовних засобів масової інформації щодо подій в Україні, з увагою до особливостей обстановки, властивій Україні останніми роками, та урахуванням лінгвістичного компоненту повідомлень, досліджених ЗМІ.

Викладення основного матеріалу

Зазвичай, основним способом моніторингу в центральних органах виконавчої влади є пошук інформації в Інтернет за напрямом діяльності за допомогою ключових слів. При цьому значний інтерес для опрацювання інформації становить аналітико-прогностичний напрям, виконання завдань за яким дозволяє сприяти управлінню державними процесами та є необхідним у діяльності інформаційно-аналітичних підрозділів державних структур [6].

Тому, в інтересах інформаційної безпеки України, необхідним є створення вітчизняних засобів моніторингу текстів іншомовних ЗМІ щодо подій в Україні та виконання аналітико-прогностичних завдань.

Основні функції системи моніторингу іншомовних ЗМІ повинні бути такими:
- збір в Інтернет іншомовних повідомлень іноземних ЗМІ про події в Україні;
- їх аналіз.

У розробці необхідна також участь фахівців-лінгвістів, які повинні володіти іноземними мовами, що є важливим у даному дослідженні, на рівні, достатньому для здійснення аналізу тексту на предмет відповідності його окремих компонентів умовам дослідження. Результати їх роботи повинні бути включені до бази знань підсистеми визначення емоційної тональності та підсистеми прогнозування реакцій аудиторії, які повинні увійти до системи моніторингу. Відповідна база знань повинна бути покладена в основу блоку Сентимент-аналізу (Sentiment-analysis) системи та безпосередньо впливати на прийняття нею рішення щодо змісту повідомлення.

Створення системи передбачає два основні етапи розробки:
- розробку лінгвістичного забезпечення;
- розробку програмного забезпечення.

Етап розробки лінгвістичного забезпечення включає аналіз природномовних текстів. Він повинен бути спрямований на формування поняттєвої структури, тобто на витяг знань з іншомовних текстів та їх прагматичну інтерпретацію в термінах поставленої задачі. При цьому кожен текст розглядатиметься як об'єкт різних рівнів аналізу [7]: як знакова система, як граматична система та як система знань про предметну галузь. Через те, що кожен рівень має свої особливості, свої засоби вираження, він передбачатиме наявність відповідних методів обробки.

Лінгвістичне розпізнавання знань з предметної галузі відображатиме морфологічний, синтаксичний та семантичний рівні мовної системи. На лінгвістичному етапі розроблення системи слід побудувати поняттєву структуру тексту. Вона включатиме тезаурус, структура якого орієнтована на завдання дослідження. Результати графемного та

синтаксичного розпізнавання є вхідними даними для семантичного розпізнавання, також як і еталонні моделі з тезаурусом понять, тезаурусом відношень і тезаурусом логіко-семантичних відношень. На етапі семантичного розпізнавання, де всі фрагменти тексту повинні бути об'єднані в єдину логіко-семантичну структуру, обробка полягатиме в узагальненні та уніфікації понять, відношень та їх характеристик [7].

Дослідження повідомлень провідних зарубіжних ЗМІ, розрахованих на англомовну аудиторію, зокрема США і країн Євросоюзу, показує наявність розбіжностей у поданні інформації про події в Україні та їх основних учасників. Зокрема, у статтях щодо конфлікту на Сході України проукраїнська позиція частіше властива ЗМІ США, а протилежна інтересам України позиція простежується часом у провідних європейських ЗМІ. Зокрема, такий розподіл часто властивий повідомленням європейських новинних каналів «Euronews» та «Deutschewelle», англомовній версії «Russia today», а також таким виданням як «Guardian», «France 24» та деяким іншим.

Важливим для удосконалення моніторингу є опрацювання лексичного наповнення повідомлень. Воно свідчить про окремі тенденції формування суспільної думки та є підґрунтям для приймання системою моніторингу рішення при виконанні завдання. Огляд вищезгаданих матеріалів показав, що їх дискурс відображає наявні там спроби впливу на аудиторію та включає такі мовні прийоми, як тенденційне викладення фактів і термінологічне мінування: викривлення первинної, правильної по суті, інформації і принципово важливих базових термінів і тлумачень.

Наприклад, подання сторін конфлікту у новинах каналу «Euronews» включає визначення: «*government forces*» («урядові війська») та *«pro-Moscow militia»* (militia – «збройні формування», «ополчення»).

Лексичний склад англомовних видань щодо подій в Україні відображає поширення наступних уявлень та стереотипів, які закладаються в основу тенденцій формування суспільної думки. Основні з них наведені в наступному переліку.

1) В Україні йде громадянська війна, однією із причин якої є розбіжності серед населення щодо укладання Угоди про асоціацію з Євросоюзом [8].

2) Збройні Сили України самі влаштовують провокації для обстрілів [9].

3) ЗС України обстрілюють житлові квартали [10].

4) Нові провладні сили підтримують українські націоналістичні угрупування ІІ Світової війни, які проводили етнічні чистки (різанина поляків), співпрацювали з нацистами та влаштовували єврейські погроми [11].

5) Населення України надто бідне, і не зможе купувати товари європейських виробників, тому для нього доцільно навести лад у своїй країні та не поспішати укладати з ЄС угоду про асоціацію та інші взаємні юридичні зобов'язання [12].

6) Воєнні дії, що ведуться на Сході України, спричиняють зловживання порушенням там прав людей і дають «карт-бланш» для порушення прав людини на території усієї країни [11].

7) Певна частина ЗС України не прагне до миру та намагається будь-яким способом спровокувати бойові дії [13].

8) Український уряд, за підтримки Заходу, придушує військами право свого населення на самовизначення [14].

9) Український уряд спекулює загрозою Російської агресії, щоб отримувати від Заходу грошову допомогу та зброю [15].

Проведене дослідження з опрацювання дискурсу зазначених повідомлень дозволяє виділити ключові блоки, в межах яких формуються теми з властивим їм простором понять. На наступному етапі поняттєву структуру тексту інтегрують до бази знань системи. Це завдання виконується шляхом процедури семантико-прагматичного розпізнавання [7].

Базу знань покладають в основу блоку Сентимент-аналізу (Sentiment-analysis), в якому використовуються відповідні алгоритми машинного навчання з інструментами класифікації. Застосування машинного навчання також допомагає у виявленні необ'єктивних або фейкових відгуків, що може бути корисним при виконанні вищевказаних завдань з моніторингу. Так, у ході аналізу тональності або вилучення поглядів проводиться математичне дослідження думок, настроїв, оцінок, тверджень, відгуків, поглядів, емоцій тощо, виражених у формі тексту. Згідно з формальним визначенням контенту, думка представлена у вигляді кортежу (e_j, a_{jk}, so_{ijkl}, h_i, t_l), де e_j – об'єкт думки, a_{jk} – аспект об'єкту e_j. А вираз so_{ijkl} – це тональність думки, яку висловлює автор h_i про аспект a_{jk} об'єкту e_j в час t_l (позитивна, негативна, нейтральна). h_i – автор думки. t_l – час, коли думка висловлюється [16].

У ході класифікації дається оцінка всьому документу в цілому. Припускається, що документ написаний одним автором про один об'єкт. Використовуються такі інструменти як «навчання без учителя» (Unsupervised classification) та «навчання зі вчителем» (Supervised classification). Вони дозволяють встановити в тексті велику кількість висловів думок та почуттів (opinion words, sentiment words), що може свідчити про суб'єктивність висловів, також як і багато прикметників. Аналіз текстів іншомовних ЗМІ проводиться з використанням алгоритмів Наївного байєсівського класифікатора та Методу опорних векторів (Supported Vector Machine). При цьому, кожен з цих методів має свої переваги та недоліки.

Так, основні переваги Наївного байєсівського класифікатора (НБК) такі:
- простота реалізації;
- швидкий процес навчання. Обчислювальна складність навчання O(|V|);
- незважаючи на те, що припущення про незалежність класифікаційних ознак не є правильним у природній мові (значення слова залежить від контексту), НБК часто показує хороші результати при класифікації текстів.

Однак, оскільки в природній мові слова не є незалежними, НБК не є оптимальним. Серед його основних недоліків є те, що значення, які повертаються при класифікації, не можна трактувати, як імовірності, що не дає можливості відповісти на питання, з якою імовірністю визначений клас [17].

Для удосконалення процесу опрацювання іншомовних ЗМІ у даному дослідженні більш доцільним вбачається використання Методу опорних векторів.

У результаті вирішення задачі класифікації текстових документів для даних, що лінійно розділяються, отримуємо функцію класифікації, яка мінімізує верхню оцінку очікуваного ризику.

Однією з проблем, які пов'язані з вирішенням задач класифікації зазначеним методом, є те, що не завжди можна знайти лінійну межу між двома класами.

У таких випадках одним із варіантів є збільшення розмірності, тобто перенос даних із площини у тривимірний простір, де можливо побудувати таку площину, яка ідеально розділить множину зразків на два класи. Опорними векторами, в такому випадку, будуть слугувати об'єкти з обох класів, які є екстремальними.

Таким чином, за допомогою додавання оператора ядра і додаткових розмірностей знаходять межі між класами у вигляді гіперплощин.

Однак, слід пам'ятати: складність побудови SVM-моделі полягає у тому, що чим вища розмірність простору, тим складніше з ним працювати. Один з варіантів роботи з даними високої розмірності – це попереднє застосування якогось методу зниження розмірності даних для виявлення найбільш істотних компонент, а потім - використання методів опорних векторів.

Однак, недоліками методу SVM є те, що для класифікації використовується не вся множина зразків, а лише мала частина, яка знаходиться на межах.

Поряд з цим, перевагою методу є те, що для класифікації саме з використанням цього методу, на відміну від більшості інших методів, достатньо невеликої кількості даних. При правильній роботі моделі, побудованій на тестовій множині, цілком можливим є застосування цього методу на реальних даних. Метод опорних векторів дозволяє:

- отримати функцію класифікації з мінімальною верхньою оцінкою ризику, що очікується;
- використовувати лінійний класифікатор для роботи з даними, що нелінійно розділюються, поєднуючи простоту з ефективністю [18].

Висновки

Отже, шляхи удосконалення підходів до моніторингу іншомовних засобів масової інформації на тему подій в Україні в сучасних умовах включають створення вітчизняних засобів моніторингу текстів іншомовних ЗМІ з можливістю виконання аналітико-прогностичних завдань.

Важливим є аналіз як змісту повідомлень, так і власне їх дискурсу. Отримання об'єктивної інформації з іншомовних ЗМІ забезпечується обов'язковим урахуванням особливостей суспільно-політичної обстановки та лінгвістичного компоненту повідомлень досліджуваних ЗМІ.

Для опрацювання інформаційних потоків, використовується Сентимент-аналіз з відповідними алгоритмами машинного навчання, зокрема застосовуються Наївна модель Байєса та Метод опорних векторів. Для виконання завдань аналітико-прогностичної спрямованості, більш оптимальним з них є метод опорних векторів. Саме він ефективно здійснює класифікацію повідомлень іншомовних ЗМІ відповідно до заданих умов. Вказані шляхи удосконалення підходів до моніторингу іншомовних ЗМІ дають можливість відповідним державним структурам забезпечити оперативне виявлення інформаційних джерел серед іншомовних ЗМІ та своєчасне реагування.

Література

1. Моніторинг засобів масової інформації. Стефанцев С.С. Всеукраїнська науково-практична конференція молодих вчених, ад'юнктів, слухачів, курсантів і студентів. «Молодіжна військова наука у Київському національному університеті імені Тараса Шевченка», 26 квітня 2016 р. К., 2016. ВІКНУ. – С.268.
2. Писарчук О.О. Технологія автоматизованої багатовимірної оперативної та поглибленої аналітичної обробки актуальних інформаційних масивів / О.О. Писарчук, В.С. Косіков // Проблеми створення, випробування, застосування та експлуатації складних інформаційних систем. - 2015. - Вип. 10. - С. 183-195. // [Електр. Ресурс]. – Режим доступу: http://nbuv.gov.ua/UJRN/Psvz_2015_10_22 .
3. Писарчук О.О. Методика прогнозування розвитку інформаційного контенту в мережі Інтернет / О. О. Писарчук, Д. В. Порада // Проблеми створення, випробування, застосування та експлуатації складних інформаційних систем. - 2015. - Вип. 10. - С. 170-182. // [Електр. Ресурс]. – Режим доступу: http://nbuv.gov.ua/UJRN/Psvz_2015_10_21 .
4. Dr. Babych B., Dr.Atwell E. Multilingual information extraction framework for real-time detection of terrorist propaganda threats in on-line communication. Військова освіта і наука: сьогодення та майбутнє: збірник матеріалів XI Міжнародної науково-практичної конференції. 27 листопада 2015 р. К. : 2015. ВІКНУ.—С.93.
5. Проблеми автоматизованого аналізу і оброблення природномовних текстів. Бойко О.В., Мірошниченко О.В., Стамбірська Р.Г. Всеукраїнська науково-практична конференція молодих вчених, ад'юнктів, слухачів, курсантів і студентів. 24 квітня 2015 р. К., 2015. ВІКНУ. — С.243.
6. Марченко-Бабіч О.М. Основні вимоги до системи моніторингу іншомовних ЗМІ в інтересах забезпечення інформаційної безпеки. Збірник наукових праць ВІ КНУ імені Тараса Шевченка. - № 51, С.119-123.
7. Бадьорина Л.М. Основи комп'ютерної лінгвістики. Навчальний посібник / Л.М.Бадьорина, І.В.Замаруєва, В.А.Широков. - К. :Видавничий центр КНУКІМ, 2011. С.107-108.
8. Won't help ordinary Ukrainians': Dutch suspect EU/Kiev deal more for power players than people // [Електр. Ресурс]. – Режим доступу: https://www.rt.com/news/338580-netherlands-ukraine-eu-association-referendum
9. Deadly battles near Donetsk signal escalation in Ukraine conflict // [Електр. Ресурс]. – Режим доступу: http://www.dw.com/en/deadly-battles-near-donetsk-signal-escalation-in-ukraine-conflict/a-18495369 .

10. Tensions in east Ukraine as Kiev ‘repels tank assault’ http://www.france24.com/en/20150811-ukraine-tensions-flare-kiev-tank-assault-Starohnativka-Novolaspa .
11. Ukraine bans Soviet symbols and criminalises sympathy for communism. // [Електр. Ресурс]. – Режим доступу: http://www.theguardian.com/world/2015/may/21/ukraine-bans-soviet-symbols-criminalises-sympathy-for-communism .
12. ‘Won’t help ordinary Ukrainians’: Dutch suspect EU/Kiev deal more for power players than people // [Електр. Ресурс]. – Режим доступу: https://www.rt.com/news/338580-netherlands-ukraine-eu-association-referendum/ .
13. Analysts: Kiev needs tensions in Donbas as excuse for belt tightening // [Електр. Ресурс]. – Режим доступу: http://tass.ru/en/opinions/835014 .
14. Kiev’s policy leads to further separation of Donbas — Russian Foreign Ministry // [Електр. Ресурс]. – Режим доступу: http://tass.ru/en/politics/820220 .
15. Kiev blames 'Russian aggression' for Ukraine's skyrocketing inflation // [Електр. Ресурс]. – Режим доступу: https://www.rt.com/business/337770-ukraine-inflation-russia-fault/ .
16. Sentiment analysis and opinion mining. Bing Liu. Dep artment of computer science. University of Illinois at Chicago. // [Електр. Ресурс]. – Режим доступу: http://www.slideshare.net/hustwj/sentiment-analysis-tutorial-aaai2011 .
17. Generative and discriminative classifiers: Naive Bayes and Logistic regression. // [Електр. Ресурс]. – Режим доступу: https://www.cs.cmu.edu/~tom/mlbook/NBayesLogReg.pdf .
18. Лекция 10: Методы классификации и прогнозирования. Метод опорных векторов. Метод "ближайшего соседа". Байесовская классификация. // [Електр. Ресурс]. – Режим доступу: http://www.intuit.ru/studies/courses/6/6/lecture/176?page=2 .

Literatura

1. Monitoring zasobiv masovoї informacії. Stefancev S.S. Vseukraїns'ka naukovo-praktichna konferencija molodih vchenih, ad’junktiv, sluhachiv, kursantiv i studentiv. “Molodizhna vijs'kova nauka u Kiїvs'komu nacional'nomu universiteti imeni Tarasa Shevchenka” 26 kvitnja 2016 r. K., 2016. VIKNU. – S.268.
2. Pisarchuk O.O. Tehnologija avtomatizovanoї bagatovimirnoї operativnoї ta pogliblenoї analitichnoї obrobki aktual'nih informacijnih masiviv / O.O. Pisarchuk, V.S. Kosikov // Problemi stvorennja, viprobuvannja, zastosuvannja ta ekspluatacії skladnih informacijnih sistem. — 2015. — Vip. 10. — S. 183-195. // [Elektr. Resurs]. – Rezhim dostupu: http://nbuv.gov.ua/UJRN/Psvz_2015_10_22
3. Pisarchuk O.O. Metodika prognozuvannja rozvitku informacijnogo kontentu v merezhi Internet / O. O. Pisarchuk, D. V. Porada // Problemi stvorennja, viprobuvannja, zastosuvannja ta ekspluatacії skladnih informacijnih sistem. — 2015. — Vip. 10. — S. 170-182. // [Elektr. Resurs]. – Rezhim dostupu: http://nbuv.gov.ua/UJRN/Psvz_2015_10_21
4. Dr. Babych B., Dr.Atwell E. Multilingual information extraction framework for real-time detection of terrorist propaganda threats in on-line communication. Vijs'kova osvita i nauka: s'ogodennja ta majbutne: zbirnik materialiv HI Mizhnarodnoї naukovo-praktichnoї konferencії. 27 listopada 2015 r. K. : 2015. VIKNU.—S.93.
5. Problemi avtomatizovanogo analizu i obroblennja prirodnomovnih tekstiv. Bojko O.V., Miroshnichenko O.V., Stambirs'ka R.G. Vseukraїns'ka naukovo-praktichna konferencija molodih vchenih, ad’junktiv, sluhachiv, kursantiv i studentiv. 24 kvitnja 2015 r. K., 2015. VIKNU. — S.243.
6. Marchenko-Babich O.M. Osnovni vimogi do sistemi monitoringu inshomovnih ZMI v interesah zabezpechennja informacijnoї bezpeki. Zbirnik naukovih prac' VI KNU imeni Tarasa Shevchenka. — № 51, S.119-123.
7. Bad'orina L.M. Osnovy komp’juternoї lingvistiki. Navchal'nij posibnik / L.M.Bad'orina, I.V.Zamaruyeva, V.A.Shirokov. — K. :Vidavnichij centr KNUKIM, 2011. S.107—108.
8. Won’t help ordinary Ukrainians’: Dutch suspect EU/Kiev deal more for power players than people // [Elektr. Resurs]. – Rezhim dostupu: https://www.rt.com/news/338580-netherlands-ukraine-eu-association-referendum
9. Deadly battles near Donetsk signal escalation in Ukraine conflict // [Elektr. Resurs]. – Rezhim dostupu: http://www.dw.com/en/deadly-battles-near-donetsk-signal-escalation-in-ukraine-conflict/a-18495369.
10. Tensions in east Ukraine as Kiev ‘repels tank assault’ // [Elektr. Resurs]. – Rezhim dostupu: http://www.france24.com/en/20150811-ukraine-tensions-flare-kiev-tank-assault-Starohnativka-Novolaspa.
11. Ukraine bans Soviet symbols and criminalises sympathy for communism. // [Elektr. Resurs]. – Rezhim dostupu: http://www.theguardian.com/world/2015/may/21/ukraine-bans-soviet-symbols-criminalises-sympathy-for-communism.

12. 'Won't help ordinary Ukrainians': Dutch suspect EU/Kiev deal more for power players than people // [Elektr. Resurs]. – Rezhim dostupu: https://www.rt.com/news/338580-netherlands-ukraine-eu-association-referendum/ .
13. Analysts: Kiev needs tensions in Donbas as excuse for belt tightening // [Електр. Ресурс]. – Режим доступу: http://tass.ru/en/opinions/835014 .
14. Kiev's policy leads to further separation of Donbas — Russian Foreign Ministry // [Elektr. Resurs]. – Rezhim dostupu: http://tass.ru/en/politics/820220 .
15. Kiev blames 'Russian aggression' for Ukraine's skyrocketing inflation // [Elektr. Resurs]. – Rezhim dostupu: https://www.rt.com/business/337770-ukraine-inflation-russia-fault/ .
16. Sentiment analysis and opinion mining. Bing Liu. Department of computer science. University of Illinois at Chicago. // [Elektr. Resurs]. – Rezhim dostupu: http://www.slideshare.net/hustwj/sentiment-analysis-tutorial-aaai2011 .
17. Generative and discriminative classifiers: Naive Bayes and Logistic regression. // [Elektr. Resurs]. – Rezhim dostupu: https://www.cs.cmu.edu/~tom/mlbook/NBayesLogReg.pdf .
18. Lekcija 10: Metody klassifikacii i prognozirovanija. Metod opornyh vektorov. Metod "blizhajshego soseda". Bajesovskaja klassifikacija. // [Elektr. Resurs]. – Rezhim dostupu: http://www.intuit.ru/studies/courses/6/6/lecture/176?page=2 .

RESUME

O.M.Marchenko-Babich

The improvement of approaches to foreign mass media monitoring on the subject of the situation in Ukraine

A large amount of foreign media news and brief terms are the main reasons to improve the approaches to information environment monitoring. Such improvement brings the need for the use of software systems to make the process of information extraction from the media and its authenticity verification more objective.

The development of foreign media monitoring concerning events in Ukraine nowadays means to create the domestic system with analytical and forecast functions applied for foreign media texts.

The analysis of messages content is so important as that for their discourse. Bothe the socio-political situation and linguistic component are the main prerequisites for getting unbiased information from foreign media. The research of vocabulary in the news about Ukraine reveals a trend of spreading stereotypes to shape public opinion.

Sentiment analysis technology with the appropriate machine learning algorithms for data flows processing include naive Bayesian classifier algorithms and Support vector machine. Support vector machine is a technique the most appropriate to be employed in the text processing, with analysis function particularly. The processing of foreign mass media with this technique provides the information and analytical activities under specified conditions. The ways of improvement of approaches to monitoring of foreign media enable relevant governmental agencies to detect a defined information activity and to respond timely.

Надійшла до редакції 27.10.2016